

深度学习 细说 PyTorch

理论、算法、模型与编程实现

凌 峰 丁麒文◎编著



清华大学出版社
北京

内 容 简 介

本书由业界专家编撰,采用理论描述加代码实践的思路,详细介绍PyTorch的理论知识及其在深度学习中的应用。全书分为两篇,共16章。第一篇为基础知识,主要介绍PyTorch的基本知识、构建开发环境、卷积网络、经典网络、模型保存和调用、网络可视化、数据加载和预处理、数据增强等内容;第二篇为高级应用,主要介绍数据分类、迁移学习、人脸检测和识别、生成对抗网络、目标检测、ViT等内容。本书内容涵盖PyTorch从入门到深度学习的各个方面,是一本基础应用与案例实操相结合的参考书。

本书理论兼备实例,深入浅出,适合PyTorch初学者使用,也可以作为理工科高等院校本科生、研究生的教学用书,还可作为相关科研工程技术人员的参考书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。举报:010-62782989, beiqinquan@tup.tsinghua.edu.cn。

图书在版编目(CIP)数据

细说PyTorch深度学习:理论、算法、模型与编程实现/凌峰,丁麒文编著. —北京:清华大学出版社, 2023.6

ISBN 978-7-302-63194-1

I. ①细… II. ①凌… ②丁… III. ①机器学习 IV. ①TP181

中国国家版本馆CIP数据核字(2023)第052569号

责任编辑:王金柱

封面设计:王翔

责任校对:闫秀华

责任印制:宋林

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦A座 邮 编:100084

社总机:010-83470000 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印装者:北京嘉实印刷有限公司

经 销:全国新华书店

开 本:185mm×235mm 印 张:20.75 字 数:498千字

版 次:2023年6月第1版 印 次:2023年6月第1次印刷

定 价:99.00元

产品编号:101495-01

前言

21世纪大国之间的竞争归根结底是人才的竞争,人工智能作为目前促进产业升级的关键技术,在学术界和工业界都有广泛的应用,被国家提高到了战略高度。广大国民掌握人工智能技术将极大地促进生产力发展,提升国家综合竞争力,为国家发展做出技术上的贡献。出于这样的思考,本书为有志于在21世纪从事人工智能事业的读者编写,希望通过本书为促进国家人工智能技术的进步做出微薄贡献。

众所周知,在国家层面,人工智能技术已经成为社会经济发展的新引擎,该技术已经被应用于人们日常生活的方方面面。PyTorch是实现人工智能的重要技术途径之一,学好PyTorch将十分有利于掌握人工智能技术。

人工智能技术作为新一轮产业变革的核心驱动技术,将进一步释放历次科技革命和产业变革积蓄的巨大势能,进一步促进经济巨大发展,形成宏观经济、微观经济等领域的社会智能化新需求,激发新技术、新产品、新产业、新业态、新模式不断涌现,引发社会经济结构重大变革,深刻改变国民日常生产生活方式和经济社会生活思维模式,实现国家生产力的跨越提升。

我国经济社会发展已经进入了新阶段,实现共同富裕已经成为当前主要的社会发展目标和任务,需要加快人工智能应用于实体经济产业领域,提升人工智能技术产业化水平,为我国社会发展注入强大的技术支持和核心科技动力。

随着我国社会发展和工业升级的需求,人工智能技术人才需求越来越大,但是由于人工智能技术是一门交叉学科,涉及的基础知识繁杂,因此不易入

门和学习。为了降低人工智能技术的入门门槛，本书从理论出发，结合实例，尽量用简单易懂的语言讲述高深的知识点，为有志于从事基于PyTorch进行人工智能开发的从业者提供一本好的PyTorch技术参考书。

本书内容

本书结合多年PyTorch使用经验与实际工程应用案例，将PyTorch的编程方法与技巧详细地讲解给读者。本书在讲解过程中步骤详尽、内容新颖，辅以相应的图示，使读者在阅读时能一目了然，从而快速掌握书中所讲的内容。

第一篇为基础知识，包括以下章节：

- | | |
|------------------|--------------|
| 第1章 人工智能和PyTorch | 第2章 开发环境 |
| 第3章 PyTorch入门 | 第4章 卷积网络 |
| 第5章 经典神经网络 | 第6章 模型的保存和调用 |
| 第7章 网络可视化 | 第8章 数据加载和预处理 |
| 第9章 数据增强 | |

第二篇为高级应用，包括以下章节：

- | | |
|--------------|-------------|
| 第10章 图像分类 | 第11章 迁移学习 |
| 第12章 人脸检测和识别 | 第13章 生成对抗网络 |
| 第14章 目标检测 | 第15章 图像风格迁移 |
| 第16章 ViT | |

本书特点

本书由资深业界专家精心编写，内容涵盖PyTorch的基础知识、经典算法、模型训练及编程实现。

从PyTorch的安装与基本知识开始，首先介绍和深度学习相关的库NumPy和Matplotlib、Scikit-Learn，然后逐步深入细致地讲解各个知识点，确保读者可以快速上手。

基础理论结合热点应用，比如，介绍了PyTorch在经典的神经网络、卷积网络、模型调用和保存、数据可视化、数据增强等方面的编程实现，介绍了PyTorch在人脸识别、生成对抗网络、图像分类、目标检测、迁移学习中的应用以及前沿技术热点ViT等。

每个知识点在讲解的过程中，配套大量示例，全书涉及100多个编程实例，向读者展示PyTorch在深度学习中的应用。

读者对象

本书适合PyTorch初学者和期望应用PyTorch技术进行机器学习开发的读者，具体说明如下：

- 初学PyTorch的技术人员
- 广大从事深度学习的科研工作人员
- 大、中专院校的教师和学生
- 相关培训机构的教师和学员
- 刚参加工作实习的深度学习“菜鸟”
- PyTorch深度学习技术爱好者

配书源码

本书提供了程序源代码，读者可扫描下面的二维码，按扫描后的页面提示填写你的邮箱，把下载链接转发到邮箱中下载。如果下载有问题或阅读中发现问题，请用电子邮件联系booksaga@126.com，邮件主题写“细说PyTorch深度学习：理论、算法、模型与编程实现”。



读者服务

为了方便解决本书的疑难问题，读者朋友在学习过程中遇到与本书有关的技术问题，可以关注“算法仿真”公众号获取帮助，我们将竭诚为您服务。

本书由凌峰、丁麒文编著，虽然作者在本书的编写过程中力求叙述准确、完善，但由于水平有限，书中疏漏之处在所难免，希望广大读者和同仁及时指出，共同促进本书质量的提高。

最后，再次希望本书能为读者的学习和工作提供帮助！

编者

2023年3月

目 录

第 1 篇 基础知识

第1章 人工智能和PyTorch	2
1.1 人工智能和深度学习	2
1.1.1 人工智能	2
1.1.2 深度学习	3
1.2 深度学习框架	5
1.3 PyTorch	7
1.3.1 PyTorch简介	7
1.3.2 PyTorch的应用领域	9
1.3.3 PyTorch的应用前景	10
1.4 小结	12
第2章 开发环境	13
2.1 PyTorch的安装	13
2.2 NumPy	16
2.2.1 NumPy的安装与查看	17
2.2.2 NumPy对象	17
2.2.3 数组	21
2.2.4 数学计算	26
2.3 Matplotlib	32
2.3.1 Matplotlib的安装和简介	33
2.3.2 Matplotlib Figure图形对象	35
2.4 Scikit-Learn	47
2.5 小结	48

第3章 PyTorch入门	49
3.1 PyTorch的模块	49
3.1.1 主要模块	49
3.1.2 辅助模块	53
3.2 张量	54
3.2.1 张量的数据类型	55
3.2.2 创建张量	56
3.2.3 张量存储	61
3.2.4 维度操作	63
3.2.5 索引和切片	65
3.2.6 张量运算	67
3.3 torch.nn模块	76
3.3.1 卷积层	76
3.3.2 池化层	80
3.3.3 激活层	87
3.3.4 全连接层	91
3.4 自动求导	92
3.5 小结	95
第4章 卷积网络	96
4.1 卷积网络的原理	96
4.1.1 卷积运算	96
4.1.2 卷积网络与深度学习	98
4.2 NumPy建立神经网络	99
4.3 PyTorch建立神经网络	101
4.3.1 建立两层神经网络	101
4.3.2 神经网络参数更新	102
4.3.3 自定义PyTorch的nn模块	103
4.3.4 权重共享	105
4.4 全连接网络	107
4.5 小结	111
第5章 经典神经网络	112
5.1 VGGNet	112
5.1.1 VGGNet的结构	112
5.1.2 实现过程	114
5.1.3 VGGNet的特点	115
5.1.4 查看PyTorch网络结构	116

5.2 ResNet	118
5.2.1 ResNet的结构	118
5.2.2 残差模块的实现	120
5.2.3 ResNet的实现	122
5.2.4 ResNet要解决的问题	126
5.3 XceptionNet	128
5.3.1 XceptionNet的结构	128
5.3.2 XceptionNet的实现	131
5.4 小结	135
第6章 模型的保存和调用	136
6.1 字典状态 (state_dict)	136
6.2 保存和加载模型	138
6.2.1 使用state_dict加载模型	138
6.2.2 保存和加载完整模型	139
6.2.3 保存和加载Checkpoint用于推理、继续训练	139
6.3 一个文件保存多个模型	140
6.4 通过设备保存和加载模型	141
6.5 小结	143
第7章 网络可视化	144
7.1 HiddenLayer可视化	144
7.2 PyTorchViz可视化	146
7.3 TensorboardX可视化	149
7.3.1 简介和安装	149
7.3.2 使用TensorboardX	150
7.3.3 添加数字	151
7.3.4 添加图片	152
7.3.5 添加直方图	153
7.3.6 添加嵌入向量	154
7.4 小结	156
第8章 数据加载和预处理	157
8.1 加载PyTorch库数据集	157
8.2 加载自定义数据集	159
8.2.1 下载并查看数据集	159
8.2.2 定义数据集类	161
8.3 预处理	164
8.4 小结	168

第9章 数据增强	169
9.1 数据增强的概念	169
9.1.1 常见的数据增强方法	170
9.1.2 常用的数据增强库	171
9.2 数据增强的实现	172
9.2.1 中心裁剪	173
9.2.2 随机裁剪	174
9.2.3 缩放	175
9.2.4 水平翻转	176
9.2.5 垂直翻转	177
9.2.6 随机角度旋转	178
9.2.7 色度、亮度、饱和度、对比度的变化	179
9.2.8 随机灰度化	180
9.2.9 将图形加上padding	181
9.2.10 指定区域擦除	182
9.2.11 伽马变换	183
9.3 小结	184

第 2 篇 高级应用

第10章 图像分类	186
10.1 CIFAR10数据分类	186
10.1.1 定义网络训练数据	187
10.1.2 验证训练结果	192
10.2 数据集划分	193
10.3 猫狗分类实战	195
10.3.1 猫狗数据预处理	195
10.3.2 建立网络猫狗分类	196
10.4 小结	199
第11章 迁移学习	200
11.1 定义和方法	200
11.2 蚂蚁和蜜蜂分类实战	202
11.2.1 加载数据	202
11.2.2 定义训练方法	204

11.2.3 可视化预测结果	205
11.2.4 迁移学习方法一：微调网络	206
11.2.5 迁移学习方法二：特征提取器	208
11.3 小结	209
第12章 人脸检测和识别	210
12.1 人脸检测	210
12.1.1 定义和研究现状	210
12.1.2 经典算法	213
12.1.3 应用领域	216
12.2 人脸识别	217
12.2.1 定义和研究现状	217
12.2.2 经典算法	220
12.2.3 应用领域	221
12.3 人脸检测与识别实战	222
12.3.1 Dlib人脸检测	222
12.3.2 基于MTCNN的人脸识别	225
12.4 小结	227
第13章 生成对抗网络	228
13.1 生成对抗网络简介	228
13.2 数学模型	230
13.3 生成手写体数字图片实战	233
13.3.1 基本网络结构	233
13.3.2 准备数据	234
13.3.3 定义网络和训练	235
13.3.4 生成结果分析	237
13.4 生成人像图片实战	238
13.4.1 DCGAN简介	239
13.4.2 数据准备	239
13.4.3 生成对抗网络的实现	241
13.5 小结	250
第14章 目标检测	251
14.1 目标检测概述	251
14.1.1 传统目标检测算法的研究现状	252
14.1.2 深度学习目标检测算法的研究现状	252

14.1.3 应用领域	253
14.2 检测算法模型	253
14.2.1 传统的目标检测模型	253
14.2.2 基于深度学习的目标检测模型	255
14.3 目标检测的基本概念	259
14.3.1 IoU	259
14.3.2 NMS	261
14.4 Faster R-CNN目标检测	264
14.4.1 网络原理	265
14.4.2 实战	269
14.5 小结	273
第15章 图像风格迁移	274
15.1 风格迁移概述	274
15.2 固定风格固定内容的迁移	277
15.2.1 固定风格固定内容迁移的原理	277
15.2.2 PyTorch实现固定风格迁移	280
15.3 快速风格迁移	288
15.3.1 快速迁移模型的原理	288
15.3.2 PyTorch实现快速风格迁移	290
15.4 小结	297
第16章 ViT	298
16.1 ViT详解	298
16.1.1 Transformer模型中的Attention注意力机制	298
16.1.2 视觉Transformer模型详解	302
16.2 ViT图像分类实战	305
16.2.1 数据准备	305
16.2.2 定义ViT模型	306
16.2.3 定义工具函数	311
16.2.4 定义训练过程	314
16.2.5 运行结果	317
16.3 小结	318
参考文献	319

第

1

篇

基础知识





本章主要介绍PyTorch在人工智能（Artificial Intelligence，AI）领域的重要性，内容包括PyTorch的发展历史、应用前景和应用领域，以及快速入门PyTorch的方法，为读者建立关于PyTorch的初步概念，帮助读者了解PyTorch的基础知识，增强学习兴趣，以便读者可以快速入门PyTorch。

学习目标：

- （1）熟悉人工智能和深度学习。
- （2）熟悉PyTorch的应用领域。
- （3）熟悉PyTorch的应用前景。

1.1 人工智能和深度学习

深度学习是人工智能的一个重要分支，尤其是近十多年来随着计算机算力的提升，深度学习技术重塑了人工智能。本节从人工智能入手，带领读者熟悉深度学习的基础知识。

1.1.1 人工智能

人工智能是一个十分宽泛的概念，现代研究通常认为人工智能是计算机学科的一个分支，是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。

1956年，约翰·麦卡锡在美国达特茅斯学院主持召开了一场人工智能夏季研讨会，创建了人工智能这一科学领域，并掀起了科技界的一场革命，从此人工智能技术逐渐改变了整个世界的科技发展。

人工智能是一门极富挑战性的科学，从事这项工作的人必须懂得计算机知识、心理学和哲学。人工智能是内容十分广泛的科学，它由不同的领域组成，如机器学习、计算机视觉等，总的来说，人工智能研究的一个主要目标是使机器能够胜任一些通常需要人类智能才能完成的复杂工作。但不

同的时代、不同的人对这种复杂工作的理解是不同的。在当代，人工智能的研究应用已经深入百姓的吃、穿、住、行等日常生活中，实际应用包括指纹识别、人脸识别、虹膜识别、掌纹识别、专家系统、自动规划、智能搜索、博弈、自动程序设计、智能控制、机器人学、语言和图像理解、遗传编程等。具体的，例如滴滴打车、百度搜索、有道翻译、搜狗输入法、天猫商城、京东商城、掌上生活、火车票人脸识别系统、核酸检测系统等，其中都涉及大量人工智能技术的应用。

当今没有统一的原理或范式指导人工智能研究，在许多问题上研究者都存在争论，其中几个长久以来仍没有结论的问题是：是否应从心理或神经方面模拟人工智能，或者像鸟类生物学对于航空工程一样，人类生物学与人工智能研究是没有关系的，智能行为能否用简单的原则（如逻辑或优化）来描述，还是必须解决大量完全无关的问题等，这些问题还都没有明确的答案。这些都需要在后续继续进行研究，但是所有的这一切都不妨碍人工智能技术取得巨大成功。

1.1.2 深度学习

深度学习由浅层学习发展而来。实际上，在20世纪50年代，就已经有浅层学习的相关研究，代表性工作主要是罗森布拉特（F. Rosenblatt）基于神经感知科学提出的计算机神经网络，即感知器，在随后的10年中，浅层学习的神经网络曾经风靡一时，特别是马文·明斯基提出了著名的XOR问题和感知器线性不可分的问题。但随后进入了一段时间的冷却期。

1986年诞生了用于训练多层神经网络的真正意义上的反向传播算法，这是现在的深度学习中仍然在使用的训练算法，奠定了神经网络走向完善和应用的基础。

1989年，LeCun设计出了第一个真正意义上的卷积神经网络，用于手写数字的识别，该算法在当时的美国银行系统得到了成功应用，需要重点说明的是，这是现在被广泛使用的深度卷积神经网络的鼻祖。

在1986—1993年，神经网络的理论得到了极大的丰富和完善，但当时很多因素限制了它的大规模使用，例如工业发展水平、计算机网络水平、计算机硬件水平等。

到了20世纪90年代，可以说进入了机器学习百花齐放的年代。在1995年诞生了两种经典的算法：支持向量机（Support Vector Machine, SVM）和AdaBoost（自适应增强），此后它们纵横江湖数十载，神经网络则黯然失色，主要由于神经网络算法计算复杂度、梯度消失以及爆炸问题，在当时计算机网络和硬件计算水平有限的大环境下限制了神经网络算法的使用。

支持向量机算法代表了核（Kernel）技术的胜利，这是一种思想，通过隐式地将输入向量映射到高维空间中，使得原本非线性的问题得到很好的处理，因此在当时得到了广泛的应用，在当今若问题规模不大时支持向量机算法仍然具有广泛的应用领域。而AdaBoost算法则代表了集成学习算法的胜利，通过将一些简单的弱分类器集成起来使用，能够达到惊人的精度，该思想现在仍被使用。

当今语音领域炙手可热的LSTM（Long Short-Term Memory，长短期记忆网络）在2000年就出

现了，这让很多读者感到惊讶。LSTM在很长一段时间内一直默默无闻，直到2013年后与深度循环神经网络（Deep-Recurrent Neural Network, Deep-RNN）整合，才在语音识别上取得成功。

由于计算机的运算能力有限，多层网络训练困难，通常都是只有一层隐含层的浅层模型，虽然各种各样的浅层机器学习模型相继被提出，在理论分析和应用方面都产生了较大的影响，但是理论分析的难度和训练方法需要很多经验和技巧，随着最近邻（K-Nearest Neighbor, KNN）等算法的相继提出，浅层模型在模型理解、准确率、模型训练等方面被超越，机器学习的发展几乎处于停滞状态。

虽然真正意义上的人工神经网络诞生于20世纪80年代，反向传播算法（BP算法）、卷积神经网络（Convolutional Neural Networks, CNN）、LSTM等早就被提出，但遗憾的是神经网络在过去很长一段时间内并没有得到大规模的成功应用，在与SVM等机器学习算法的较量中处于下风。原因主要有：算法本身的问题，如梯度消失问题，导致深层网络难以训练；训练样本数的限制；计算能力的限制。直到2006年，随着计算机硬件技术的飞速发展，算力不再受限，移动网络技术飞速发展，世界互联，产生了巨量的数据，情况才慢慢改观。

神经网络研究领域领军者Hinton在2006年提出了神经网络深度学习（Deep Learning）算法，被众多学习者奉为该领域的经典著作，使神经网络的能力大大提高，向支持向量机算法发出挑战。2006年，机器学习领域的泰斗Hinton和他的学生Salakhutdinov在顶尖学术刊物《科学》（*Science*）上发表了一篇文章，开启了深度学习在学术界和工业界的浪潮，微软、IBM等工业巨头投入了大量资源开展研究，MIT等科研机构更是成果层出不穷。

Hinton的学生Yann LeCun的LeNets深度学习网络可以被广泛应用在全球的ATM机和银行之中。同时，Yann LeCun和吴恩达等认为卷积神经网络允许人工神经网络进行快速训练，因为它所占用的内存非常小，无须在图像上的每一个位置上都单独存储滤镜，因此非常适合构建可扩展的深度网络，因此卷积神经网络非常适合识别模型。这些人都是机器学习领域全球的领导者。

目前，新的深度学习算法面临的主要问题更加复杂，深度学习的应用领域从广度向深度发展，这对模型训练和应用都提出了更高的要求。随着人工智能的发展，冯·诺依曼式的有限状态机的理论基础越来越难以应对目前神经网络中层数的要求，这些都对深度学习新的算法的发展和应用提出了挑战。

深度学习的发展并不是一帆风顺的，经历了螺旋式上升的过程，机遇与困难并存。凝聚国内外大量的研究学者的成果才有了今天人工智能的空前繁荣，是量变到质变的过程，也是内因和外因的共同结果，符合客观事物发展规律。目前处于一个美好的时代，是深度学习的波峰时期。

深度学习技术通过具体的深度学习框架来实现，下一节简要介绍目前一些常见的深度学习框架。

1.2 深度学习框架

深度学习是目前人工智能的重要发展方向，在开始深度学习项目之前，选择一个合适的框架是非常重要的，因为选择一个合适的框架能起到事半功倍的效果。研究者使用各种不同的框架来达到研究目的，也从侧面印证了深度学习领域的百花齐放。

在深度学习初始阶段，每个深度学习研究者都需要写大量的重复代码，为了提高工作效率，研究者就将这些代码写成了一个框架放到网上让所有研究者一起使用，接着网上就出现了不同的框架。随着时间的推移，最为好用的几个框架被大量的人使用从而流行了起来，全世界最为流行的深度学习框架有PaddlePaddle、TensorFlow、Caffe、Theano、MXNet、Torch和PyTorch。

1. PaddlePaddle

PaddlePaddle（飞桨）以百度多年的深度学习技术研究和业务应用为基础，是中国首个自主研发、功能完备、开源开放的产业级深度学习平台，集深度学习核心训练和推理框架、基础模型库、端到端开发套件和丰富的工具组件于一体。

目前，PaddlePaddle已凝聚超过265万开发者，服务企业10万家，基于飞桨开源深度学习平台产生了34万个模型。PaddlePaddle可以助力开发者快速实现AI想法，快速上线AI业务，帮助越来越多的行业完成AI赋能，实现产业智能化升级。

PaddlePaddle在业内率先实现了动静统一的框架设计，兼顾灵活性与高性能，并提供一体化设计的高层API和基础API，确保用户可以同时享受开发的便捷性和灵活性。

在大规模分布式训练技术上，PaddlePaddle率先支持了千亿稀疏特征、万亿参数、数百节点并行训练的能力，并推出了业内首个通用异构参数服务器架构，已达到国际领先水平。

PaddlePaddle拥有强大的多端部署能力，支持云端服务器、移动端以及边缘端等不同平台设备的高速推理；PaddlePaddle推理引擎支持广泛的AI芯片，已经适配和正在适配的芯片或IP达到29款，处于业界领先地位。

PaddlePaddle还围绕企业实际研发流程量身定制打造了大规模的官方模型库，算法总数达到270多个，服务企业遍布能源、金融、工业、农业等多个领域。

2. TensorFlow

Google（谷歌）开源的TensorFlow是一款使用C++语言开发的开源数学计算软件，使用数据流图（Data Flow Graph）的形式进行计算。图中的节点代表数学运算，而图中的线条表示多维数据数组（Tensor，张量）之间的交互。

TensorFlow灵活的架构可以部署在一台或多台具有多个CPU、GPU的台式机及服务器中，或者

使用单一的API应用于移动设备中。TensorFlow最初是由研究人员和Google Brain（谷歌大脑）团队针对机器学习和深度神经网络进行研究而开发的，开源之后几乎可以在各个领域使用。

TensorFlow是全世界使用人数最多、社区最为庞大的一个框架，因为是Google公司出品的，所以维护与更新比较频繁，并且有着Python和C++的接口，教程也非常完善，同时很多论文复现的第一个版本的设计都是基于TensorFlow编写的，所以是深度学习界设计框架默认的老大。

3. Caffe

和TensorFlow名气一样大的是深度学习框架Caffe，由加州大学伯克利分校计算机科学博士获得者的贾扬清开发，全称是Convolutional Architecture for Fast Feature Embedding，是一个清晰而高效的开源深度学习框架，由伯克利视觉中心（Berkeley Vision and Learning Center, BVLC）进行维护。

Caffe对于卷积网络的支持特别好，同时也是用C++写的，提供C++接口，也提供MATLAB接口和Python接口。

Caffe之所以流行，是因为之前很多ImageNet比赛里面使用的网络都是用Caffe编写的，所以如果想使用这些比赛的网络模型就只能使用Caffe，这也就导致很多人直接转到Caffe这个框架下面。

Caffe的缺点是不够灵活，同时内存占用高，Caffe的升级版本Caffe 2已经开源了，修复了一些问题，同时工程水平得到了进一步提高。

4. Theano

Theano于2008年诞生于蒙特利尔理工学院，其派生出了大量的深度学习Python软件包，最著名的包括Blocks和Keras。Theano的核心是一个数学表达式的编译器，它了解如何获取结构，并将其转化为高效代码，以便在CPU或GPU上尽可能快地运行NumPy高效的本地库，如BLAS和本地代码（C++）。它是为深度学习中处理大型神经网络算法所需的计算而专门设计的，是这类库的首创之一（发展始于2007年），被认为是深度学习研究和开发的行业标准。

但是开发Theano的研究人员大多去了Google参与TensorFlow的开发，所以某种程度来讲，TensorFlow就像Theano的孩子。

5. MXNet

MXNet的主要作者是李沐，最早就是几个人抱着纯粹对技术和开发的热情做起来的，如今成为亚马逊的官方框架，有着非常好的分布式支持，而且性能特别好，占用显存低，同时它开发的语言接口不仅有Python和C++，还有R、MATLAB、Scala、JavaScript等，可以说能够满足使用任何语言的人。

但是MXNet的缺点也很明显，教程不够完善，使用的人不多，导致社区不大，同时每年很少有比赛和论文是基于MXNet实现的，这就使得MXNet的推广力度和知名度不高。

6. Torch

Torch是一个有大量机器学习算法支持的科学计算框架，它已经存在了10年，但真正开始流行起来是因为Facebook开源了大量Torch的深度学习模块和扩展。Torch的特点在于非常灵活，但另一个特殊之处是采用了Lua编程语言。在深度学习大部分以Python为编程语言的环境下，一个以Lua为编程语言的框架处于不小的劣势，这使得使用Torch框架的学习成本更高。

7. PyTorch

PyTorch是2017年1月FAIR（Facebook AI Research）发布的一款深度学习框架，从名称可以看出，PyTorch是由Py和Torch构成的。

PyTorch提供了两个高级功能：一是具有强大的GPU加速的张量计算（如NumPy）；二是包含自动求导系统的深度神经网络。

PyTorch的前身便是Torch，其底层和Torch框架一样，但是使用Python重新写了很多内容，不仅更加灵活，支持动态图，而且提供了Python接口。

除了Facebook外，PyTorch已经被Twitter、CMU和Salesforce等机构采用。

深度学习框架的出现降低了深度学习入门的门槛，不需要从复杂的神经网络开始编代码，可以根据需要选择已有的模型，通过训练得到模型参数，也可以在已有模型的基础上增加自己的神经网络层（Layer），或者是在顶端选择自己需要的分类器和优化算法（比如常用的梯度下降法）。当然正因如此，没有什么框架是完美的，就像一套积木里可能没有需要的那一种积木，所以不同的框架适用的领域不完全一致。总的来说，深度学习框架提供了一系列的深度学习组件（包括通用算法的实现），当需要使用新的算法时，用户需要自行定义并调用深度学习框架提供的函数接口来使用用户自定义的新算法。

经过这些年的发展，PyTorch受到越来越多从业者的使用和青睐，目前学术界大量使用该框架，参与这些学术研究的学生毕业之后将逐渐进入产业界，因此PyTorch必定在将来得到越来越广泛的应用。

1.3 PyTorch

本节介绍 PyTorch 及其应用领域。

1.3.1 PyTorch 简介

PyTorch是一个基于Torch的Python开源机器学习库，用于自然语言处理等应用程序。它主要由Facebook的人工智能小组开发，不仅能够实现强大的GPU加速，同时还支持动态神经网络，这一点是现在很多主流框架（如TensorFlow）都不支持的。

TensorFlow和Caffe都是命令式的编程语言，而且是静态的，首先必须构建一个神经网络，然后一次又一次使用相同的结构，如果想要改变网络的结构，就必须从头开始。但是对于PyTorch，通过反向求导技术，可以零延迟地任意改变神经网络的行为，而且实现速度快。正是这一灵活性，是PyTorch对比TensorFlow的最大优势。

另外，对比TensorFlow，PyTorch的代码更加简洁直观，底层代码也更容易看懂，这对于使用它的人来说理解底层肯定是一件令人激动的事。

PyTorch主要经历了以下发展阶段：

- 2017年1月正式发布PyTorch。
- 2018年4月更新0.4.0版，支持Windows系统，Caffe 2正式并入PyTorch。
- 2018年11月更新1.0稳定版，是GitHub增长第二快的开源项目。
- 2019年5月更新1.1.0版，支持TensorBoard，增强可视化功能。
- 2019年8月更新1.2.0版，更新torchvision、torchaudio和torchtext，增加了更多功能。

PyTorch具有以下优点：

- PyTorch是相当简洁且高效快速的框架。
- 设计追求最少的封装。
- 设计符合人类思维，它让用户尽可能地专注于实现自己的想法。
- 与Google的TensorFlow类似，FAIR的支持足以确保PyTorch获得持续的开发更新。
- PyTorch作者亲自维护的论坛供用户交流和求教问题。
- 讨论社区发达，学习材料丰富。
- 学术界和工业界广泛使用，从而强者愈强。
- 入门简单。

当然，现今任何一个深度学习框架都有其缺点，PyTorch也不例外，针对移动端、嵌入式部署以及高性能服务器端的部署其性能表现有待提升。

PyTorch现在已经从Meta“独立”出来了。2022年9月，扎克伯格亲自宣布，PyTorch基金会已新鲜成立，并归入Linux基金会旗下、其管理委员会成员包括Meta、AMD、AWS、谷歌云、微软和英伟达。

Meta表示：PyTorch成功背后的驱动力是开源社区充满活力的持续增长，成立基金会将确保在今后的许多年中，社区成员以透明和公开的方式做出决定。

成立PyTorch基金会核心就是两个字：中立。而其优先任务是确保PyTorch商业化和技术治理之间的相互独立。

Linux基金会进一步对PyTorch“中立”的重要性进行了解释：PyTorch最初由Meta的AI团队孵化，现在已经发展成为一个由贡献者和用户组成的庞大社区。在AI/ML（人工智能/机器学习）领

域，PyTorch恰如一把瑞士军刀——AI/ML社区中的大量技术都是基于PyTorch构建的。

截至2022年8月，PyTorch已经和Linux内核、Kubernetes等并列成为世界上增长最快的5个开源社区之一。从2021年8月到2022年8月，PyTorch统计了超过65000次提交，有超过2400个贡献者参与其中。对于这样的关键技术基础平台而言，中立性和真正归属于社区的特性将加速其增长，并使之更加成熟。PyTorch基金会加入Linux基金会旗下使社区成员相信，PyTorch是可以永久依赖和信任的公共资源的一部分。

1.3.2 PyTorch 的应用领域

PyTorch应用广泛，无论是在军事领域还是民用领域，都有PyTorch施展的机会，主要包括以下几个方面。

1. 数据分析与挖掘

数据挖掘和数据分析通常被相提并论，并在许多场合被认为是可以相互替代的术语。关于数据挖掘，已有多种文字不同但含义接近的定义，例如识别出巨量数据中有效的、新颖的、潜在有用的最终可理解的模式的非平凡过程，无论是数据分析还是数据挖掘，都是帮助人们收集、分析数据，使之成为信息，并做出判断，因此可以将这两项合称为数据分析与挖掘。

数据分析与挖掘技术是机器学习算法和数据存取技术的结合，利用机器学习提供的统计分析、知识发现等手段分析海量数据，同时利用数据存取机制实现数据的高效读写。机器学习在数据分析与挖掘领域拥有不可取代的地位，2012年Hadoop进军机器学习领域就是一个很好的例子。

2. 模式识别

模式识别起源于工程领域，而机器学习起源于计算机科学，这两个不同学科的结合带来了模式识别领域的调整和发展。模式识别研究主要集中在两个方面：

- (1) 研究生物体（包括人）是如何感知对象的，属于认识科学的范畴。
- (2) 在给定的任务下，如何用计算机实现模式识别的理论和方法，这些是机器学习的长项，也是机器学习研究的内容之一。

模式识别的应用领域广泛，包括计算机视觉、医学图像分析、光学文字识别、自然语言处理、语音识别、手写识别、生物特征识别、文件分类、搜索引擎等，而这些领域正是机器学习大展身手的舞台，因此模式识别与机器学习的关系越来越密切。

3. 生物信息学

随着基因组和其他测序项目的不断发展，生物信息学研究的重点正逐步从积累数据转移到如何解释这些数据。机器学习的强大学习能力和推理能力已经被用在生物信息学。在未来，生物学的

新发现将极大地依赖于在多个维度和不同尺度下对多样化的数据进行组合和关联的分析能力,而不再仅依赖于对传统领域的继续关注。

序列数据 (Sequence Data) 将与生物信息 (包括结构和功能数据、基因表达数据、生化反应通路数据、表现型和临床数据等一系列数据) 相互集成。如此庞大的数据量,在生物信息的存储、获取、处理、浏览和可视化等方面,都对理论算法和软件的发展提出了迫切的需求。

另外,由于基因组数据本身的复杂性,也对理论算法和软件的发展提出了迫切的需求。而机器学习方法,如神经网络、遗传算法、决策树和支持向量机等正适合处理这种数据量大、含有噪声并且缺乏统一理论的领域。例如,目前有大量关于新冠肺炎的机器学习论文发表。

4. 其他领域

国内外的IT巨头正在深入研究和应用机器学习,这些巨头把目标定位于全面模仿人类大脑,试图创造出拥有人类智慧的机器大脑。另外,还有一些深入日常生活的具体应用。

(1) 虚拟助手。顾名思义,当使用语音发出指令后,它们会协助查找信息。对于回答,虚拟助手会查找信息,回忆相关查询,或向其他资源(如电话应用程序)发送命令以收集信息,甚至可以指导助手执行某些任务,例如设置7点的闹钟等。

(2) 交通预测。生活中经常使用GPS导航服务,当这样做时,当前的位置和速度被保存在中央服务器上进行管理,之后使用这些数据用于构建当前流量的映射。通过机器学习可以解决配备GPS的汽车数量较少的问题,在这种情况下,机器学习有助于根据估计找到拥挤的区域。

(3) 过滤垃圾邮件和恶意软件。电子邮件客户端使用了许多垃圾邮件过滤方法,为了确保这些垃圾邮件过滤器能够不断更新,使用了机器学习技术。多层感知器和决策树归纳等是由机器学习提供支持的一些垃圾邮件过滤技术,每天可检测到超过400 000个恶意软件,每个代码与之前版本有90%~98%相似,由机器学习驱动的系统安全程序理解编码模式。因此,可以轻松检测到2%~10%变异的新恶意软件,并提供针对它们的保护。

(4) 快速揭示细胞内部结构。借由高功率显微镜和机器学习,科学家们可查看各种新冠肺炎病毒的变种,并使用机器学习方法模拟病毒变异和传播的规律,为人类健康做出贡献。

1.3.3 PyTorch 的应用前景

值得一提的是,有统计数据显示,现在NeurIPS、ICML等机器学习顶会中,有超过80%研究人员使用的都是PyTorch。

随着DeepMind的AlphaGo在2016年战胜了围棋世界冠军李世石,人工智能这个词开始进入大众的视野。从那时起,无论是大型互联网公司还是初创企业都开始大规模招聘机器学习的相关从业者,无论是社招的求职者还是校招的应聘学生都出现了大规模的增长。

1. 计算机视觉

计算机视觉（Computer Vision, CV）方向无论是在学校还是在公司，都有着大量的从业者，并且ImageNet项目可以提供上千万的标注图片供相关从业者使用。既然ImageNet是开源的数据集，那么无论是学校的教授还是学生，无论是大型互联网公司还是初创企业，都可以轻易地获取这些数据集，不仅可以进行CV算法的研究工作，还可以进行相关的工程实践。

由于计算机视觉方向历史悠久，计算机系、工程系甚至数学系都有着大量的老师和相应的学生从事该方向的研究工作，因此学校或者研究所对工业界输出的计算机视觉人才数量也是可观的。

2. 自然语言处理

与计算机视觉相比，自然语言处理（Natural Language Processing, NLP）在自动驾驶的车机系统、自动翻译系统、人工助手系统等领域都有广阔的应用前景。

3. 推荐系统

机器学习为客户推荐引擎提供了动力，增强了客户体验，并能提供个性化体验。在这种场景中，算法处理单个客户的数据点，比如客户过去的购买记录、公司当前的库存、其他客户的购买历史等，来确定适合向每个客户推荐的产品和服务。

大型电子商务公司使用推荐引擎来增强个性化并加强购物体验，这种机器学习应用程序的另一个常见应用是流媒体娱乐服务，它使用客户的观看历史、具有类似兴趣客户的观看历史、有关个人节目的信息和其他数据点，向客户提供个性化的推荐，在线视频平台则使用推荐引擎技术帮助用户快速找到适合自己的视频。

4. 客户流失评估

企业使用人工智能和机器学习可以预测客户关系何时开始恶化，并找到解决办法。通过这种方式，新型机器学习能帮助公司处理最古老的业务问题：客户流失。

在这里，算法从大量的历史、人数统计和销售数据中找出规律，确定和理解为什么一家公司会失去客户。然后，公司就可以利用机器学习能力来分析现有客户的行为，以提醒业务人员哪些客户面临着将业务转移到别处的风险，从而找出这些客户离开的原因，然后决定公司应该采取什么措施留住客户。

流失率对于任何企业来说都是一个关键的绩效指标，对于订阅型和服务型企业来说尤为重要，例如媒体公司、音乐和电影流媒体公司、软件即服务公司以及电信公司等技术主要适用行业都面临着高流失率的压力。

5. 欺诈检测

机器学习理解模式的能力，以及立即发现模式之外异常情况的能力使它成为检测欺诈活动的宝贵工具。事实上，金融机构多年来一直在这个领域使用机器学习。

它的工作原理是这样的：数据科学家利用机器学习来了解单个客户的典型行为，比如客户在何时何地使用信用卡。机器学习可以利用这些信息以及其他数据集，在短短几毫秒内准确判断哪些交易属于正常范围，因此是合法的，而哪些交易超出了预期的规范标准，因此可能是欺诈的。机器学习在各行业中检测欺诈的应用包括金融服务、旅行、游戏和零售等。

6. 自动驾驶

自动驾驶软件技术主要分为计算机视觉、行为预测以及路径规划。很多人误以为，自动驾驶面临最难攻克的技术在于计算机视觉，但实际上并非如此。过去数年里，随着深度学习的广泛应用，计算机视觉技术发展迅速，只要经过足够的训练和提供充足的数据，计算机视觉就可以探测大部分的情景。如今，自动驾驶面临的难题主要集中在行为预测和路径规划方面。

如果回忆小时候是如何学会骑自行车的，会发现其实大人们并不会告诉孩子到底该怎么骑，主要还是靠自己探索，最终习惯了也就熟练了。

近年来，一些自动驾驶公司开始探索这种方式，通过利用有限的人类提供的数据“主动学习”如何驾驶。这就需要将大量的机器学习运用于行为预测以及路径规划。例如特斯拉公司的L2自动驾驶系统、丰田公司的THS L2自动驾驶系统。

目前深度学习已经应用于人们日常生活的各个方面，如智能手机、购物网站、旅游网站、导航地图、支付系统等，虽然有些人不知道这些领域已经应用了深度学习算法，但是深度学习确实已经深入老百姓的日常生活。学习基于PyTorch的深度学习技术必将有一个广阔的就业前景。

1.4 小结

本章讲解了人工智能、深度学习以及PyTorch的概念、应用领域和应用前景，读者通过本章的学习可以对PyTorch有一个大致的了解，引起读者的学习兴趣，PyTorch的相关内容将在后续章节逐一展开学习。